When temporal prediction errs

Neuropsychologia 134 (2019) 107200

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

When temporal prediction errs: ERP responses to delayed action-feedback onset

Ana P. Pinheiro [a,b,*], Michael Schwartze [b], Francisco Gutierrez [a], Sonja A. Kotz [b,c]

[a] CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal
[b] Faculty of Psychology and Neuroscience, University of Maastricht, Maastricht, the Netherlands
[c] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Keywords:
Temporal prediction
Sensory attenuation
Feedback delay
Prediction error
Internal forward models
ERP

ABSTRACT

Sensory suppression effects observed in electroencephalography (EEG) index successful predictions of the type and timing of self-generated sensory feedback. However, it is unclear how precise the timing prediction of sensory feedback is, and how temporal delays between an action and its sensory feedback affect perception. The current study investigated how prediction errors induced by delaying tone onset times affect the processing of sensory feedback in audition. Participants listened to self-generated (via button press) or externally generated tones. Self-generated tones were presented either without or with various delays (50, 100, or 250 ms; in 30% of trials).

Comparing listening to externally generated and self-generated tones resulted in action-related P50 amplitude suppression to tones presented immediately or 100 ms after the button press. Subsequent ERP responses became more sensitive to the type of delay. Whereas the comparison of actual and predicted sensory feedback (N1) tolerated temporal uncertainty up to 100 ms, P2 suppression was modulated by delay in a graded manner: suppression decreased with an increase in sensory feedback delay. Self-generated tones occurring 250 ms after the button press additionally elicited an enhanced N2 response.

These findings suggest functionally dissociable processes within the forward model that are affected by the timing of sensory feedback to self-action: relative tolerance of temporal delay in the P50 and N1, confirming previous results, but increased sensitivity in the P2. Further, they indicate that temporal prediction errors are treated differently by the auditory system: only delays that occurred after a temporal integration window (~100 ms) impact the conscious detection of altered sensory feedback.

1. Introduction

Life in a complex sensory world requires rapid distinctions between sensory input produced by one's own actions and input resulting from the actions of others. To facilitate this process, an internal forward model is thought to guide perception on the basis of predictions about the sensory feedback to self-action (Friston, 2005; Schroger et al., 2015). An efference copy (Holst and Mittelstaedt, 1950; von Holst, 1954; see also Sommer and Wurtz, 2008 for a review) of the motor command is sent from the motor cortex to relevant sensory regions via the cerebellum, supporting the computation of the predicted sensory feedback to self-action (Knolle et al., 2013a, 2012; Wolpert and Flanagan, 2001).

An efficient perceptual system dissociates processing of externally generated sensory input from self-generated sensory feedback as the latter is typically highly predictable and does not signal novelty. Electroencephalographic (EEG) neuroimaging studies show that activity in sensory brain regions is suppressed in response to sensory feedback to self-action compared to similar input of external origin (e.g., Baess et al., 2008; Knolle et al., 2013a). However, when sensory feedback does not match the predicted sensation, a prediction error signal (i.e., surprise – Friston, 2010) is generated, which translates into increased activity in the sensory cortex (e.g., Behroozmand et al., 2009; Knolle et al., 2013b).

In event-related potential (ERP) studies, amplitude attenuation in response to self-generated sensory input (compared to physically identical but externally generated [i.e., less predictable] input) has been observed in the N1 component (e.g., Baess et al., 2008; Ford et al., 2001b; Knolle et al., 2013b, 2013a). Although less consistent, such attenuation is also found in the P2 component (Knolle et al., 2012,

* Corresponding author.
E-mail address: appinheiro@psicologia.ulisboa.pt (A.P. Pinheiro).

https://doi.org/10.1016/j.neuropsychologia.2019.107200
Received 20 January 2019; Received in revised form 18 September 2019; Accepted 19 September 2019
Available online 23 September 2019

2013a; 2013b). The auditory N1 (peaking approximately 100 ms [ms] post-stimulus onset) is generated in the primary (Heschl's gyrus) and secondary (planum temporale) auditory cortices (e.g., Godey et al., 2001; Näätänen and Michie, 1979; Zouridakis et al., 1998), but may also receive contributions from frontal brain regions (Näätänen and Picton, 1987). In speech production[1] (e.g., Ford et al., 2001a; Perez et al., 2012; Wang et al., 2014) or button press[2] tasks (e.g., Baess et al., 2011; Knolle et al., 2012, 2013a), N1 amplitude modulation is attributed to the comparison between predicted and perceived sensory feedback: N1 amplitude is reduced in response to sounds elicited by one's own action compared to passive listening to the same sounds. However, when auditory feedback does not match the predicted sensation (prediction error), N1 amplitude is increased (Behroozmand et al., 2009; Behroozmand and Larson, 2011; Knolle et al., 2013b).

Multiple neural sources contribute to the P2, including the planum temporale and auditory association areas (Godey et al., 2001). Traditionally, the P2 (peaking approximately 200 ms post-stimulus onset) has been associated with attention and categorization processes (reviewed in Crowley and Colrain, 2004), even though its functional significance is less well understood. Studies probing action-sound contingencies have interpreted P2 attenuation as a reflection of the conscious detection of a self-generated sound (Knolle et al., 2013b, 2013a, 2012). When expected self-generated sensory feedback is modified, suppression is reduced (e.g., Behroozmand et al., 2011; Knolle et al., 2013b). Differences in the processing of self-generated and externally generated input have also been observed in middle latency responses preceding the N1 (i.e., approximately at 27–33 ms after sound onset; Baess et al., 2009), which are sensitive to the temporal regularity of a stimulus (Gorina-Careta et al., 2016; Leung et al., 2013). The P50 (even though not directly examined in the context of button press tasks) was also found to be sensitive to temporal stimulus predictability (Schwartze et al., 2013; White and Yee, 2006).

Effects of unexpected sensory feedback (indicating a prediction error) were also manifested in components later than the P2: when sensory feedback is modified (comparing self-generated with externally generated tones including a 30% pitch change), enhanced N2 and P3a responses to the unexpected sounds seem to relate to the saliency of a prediction error (Knolle et al., 2013b).

1.1. Effects of temporal predictability on sensory attenuation

In a dynamically changing environment, acoustic features and the timing of sounds can be unpredictable even when these sounds are generated by one's own actions. For example, when making a phone call from a remote hotel with an inefficient telecommunication system, the sensory feedback to our voice may be delayed. Efficient anticipation of the sensory consequences of one's own actions requires not only a precise prediction regarding the type of information conveyed by sensory feedback (*what*) but also a precise prediction of *when* this feedback will occur (Schwartze et al., 2013; Schwartze and Kotz, 2013). In auditory perception, temporal (*when*) and identity (*what*) predictions may be subserved by different neurofunctional mechanisms (Hsu et al., 2013) and modulate sensory processing in a different manner (Hsu et al., 2013). Precise predictions about both type and timing of sensory feedback may result in stronger suppression effects (Baess et al., 2008). However, the effects of temporal (*when*) prediction might occur at earlier stages (reflected in modulations of the N1 and P2 ERP components) than the effects of content (*what*) prediction (reflected in modulations of the P3 ERP component) (Hsu et al., 2013).

Self-initiation provides a strong temporal cue for the sensory feedback elicited by an action. There is limited research on how temporal manipulations of action-sound contingencies affect the operations of the internal forward model (Table 1). The existing studies using a button press task with fixed or variable action-feedback delays demonstrated that sounds occurring shortly after the action tend to be associated with smaller N1 amplitudes than sounds occurring with a delay after an action (see Table 1) but only if the sound frequency was also predictable (Baess et al., 2008). A linear decrease in the N1 suppression was reported with increasing fixed (i.e., predictable) delays (from 0 ms to 100 ms with 25 ms delay increments – Oestreich et al., 2016); 0 ms *vs.* 50 ms and 100 ms delay – Whitford et al. (2011) between a button press and a tone. However, this modulation was not observed for externally generated tones with the same delays. Only one study has directly examined suppression effects beyond the N1, reporting that the P2 amplitude increased the larger the delay between the action and its sensory feedback was (Timm et al., 2016). This finding agrees with the observation that the P2 is more sensitive to the effects of temporal prediction (Sowman et al., 2012). Of note, action-feedback delays were associated with a reduced sense of agency, which affected the P2 but not the N1 component (Timm et al., 2016).

Other studies (not included in Table 1) revealed that an action-related suppression of the N1 and P2 in response to self-initiated tones may occur independently of whether a tone onset is predictable or not. In other words, sensory suppression may not depend on the contingent relations between actions and their corresponding sensory feedback (Horváth, 2015), but rather on the temporal proximity between an action and a sound (Horváth et al., 2012). For example, Horváth et al. (2012) found that tones (with random onset-to-onset intervals) coinciding with button presses also resulted in suppressed N1 and P2 amplitudes (but see Knolle et al., 2013a for negative findings). Furthermore, Baess and collaborators (2008) reported N1 suppression in response to self-initiated tones compared to externally generated tones even when the onset of sensory feedback was unpredictable; indeed, N1 suppression was largest in the case of unpredictable compared to predictable tone onset.

Despite the fact that sensory suppression could still be observed with delays up to 4 s between an action and its sensory feedback (the onset of self-initiated sounds occurred with a fixed, thus predictable, delay: Schafer and Marcus, 1973), studies with tactile feedback (Blakemore et al., 1999) or auditory feedback in active vocal production (Behroozmand and Larson, 2011; Tian and Poeppel, 2015) indicate that action-related suppression occurs only when sensory feedback is presented within 100–200 ms after the action. For example, a study of tactile stimulation, in which the sensation was produced either immediately, or 100 ms, 200 ms, or 300 ms after participants moved a robotic hand to produce tactile stimulation to their right hand, showed that salience ratings of the self-generated stimulation increased proportionally with increasing action-feedback delays: for the longest delay (300 ms), the salience of the stimulation was no different from the condition in which the robotic hand's movements were externally generated (Blakemore et al., 1999). A time-constrained window in action-related sensory suppression is consistent with the proposal that prediction and sensory feedback integrate within 200 ms after stimulus onset (see Tian and Poeppel, 2015 for evidence supporting the existence of temporal integration windows in sensorimotor processing). Temporal integration is observed when two sequential stimuli occur in the same temporal window in sensory processing and are combined to form a unitary event (e.g., Mates et al., 1994; Näätänen et al., 2007). Accordingly, perceptual changes beyond the temporal threshold (200 ms) may be treated as externally-generated.

[1] In speech production paradigms, brain activity in response to listening to self-generated speech (e.g., vocalizing/a/) is compared to brain activity in response to passively listening to pre-recorded speech.

[2] In a button-press paradigm, brain activity in response to a sound elicited by the action of pressing a button is compared to brain activity in response to an externally generated sound.

Table 1
A review of button-press studies probing the effects of modified temporal expectations on the N1 suppression to self-generated sounds.

	Sample	Stimulus	Delay	Fixed *vs.* variable delay	Task	Dependent Variable	Results
Baess et al. (2008)	n = 16 (Age: 18–31)	Sinusoidal tone •50 ms •85 dB •1000 Hz (predictable frequency) or 400–1990 Hz (unpredictable frequency)	500–1000 ms	Variable (unpredictable)	AMC; AOC; MOC	N1: 102–142 ms (PA)	AMC: UND < DEL; AOC: UND = DEL; Suppression: UND > DEL
Timm et al. (2016)	n = 17 (Age: 18–28)	Sinusoidal tone •50 ms •1000 Hz	200 ms	Fixed	AMC[a]; AOC[a]; MOC	N1: 80–100 ms (MA) P2: 150–205 ms (MA)	AMC < AOC irrespective of delay AMC = AOC; DEL > UND
Elijah et al. (2016)	*Delayed training* n = 25 *Immediate training* n = 25 (Age: 18–32)	Sinusoidal tone •100 ms •500 Hz •70 dB	100 ms	Fixed	AMC; AOC; MOC	N1: 75–125 ms (MA)	*Delayed training:* 0 = 100 ms post-training; decrease across training blocks *Immediate training:* 0 < 100 ms at pre- and post-training; no change over time
Oestreich et al. (2016)	*Low schizotypy* n = 41 (Age: 18–36) *High schizotypy* n = 39 (Age: 18–32)	Sinusoidal tone •100 ms •500 Hz	25 *vs.* 50 *vs.* 75 *vs.* 100 ms	Fixed	AMC; AOC; MOC	N1: 75–125 ms (PA)	*Low schizotypy:* Suppression: 0 = 25; 0 > 50; 25 > 50 ms
Whitford et al. (2011)	*SZ* n = 21 *HC* n = 25 (Age: 25–72)	Syllable "ah" •< 300 ms •< 85 dB	0 *vs.* 50 *vs.* 100 ms	Fixed	AMC; AOC; MOC	N1: 50–175 ms (PA)	*HC:* Suppression: 0 > 50; 50 = 100 ms *SZ:* Suppression: 50 > 0; 0 = 100 ms

Note: AMC = tone triggered by the participant's button press; AOC = externally presented tone; UND = undelayed; DEL = delayed; MA = mean amplitude; PA = peak amplitude; SZ = schizophrenia; HC = healthy controls.
The literature review was focused on button press tasks and excluded tasks comparing vocalizing with passive listening to pre-recorded vocalizations. This was motivated by the fact that differences in sound perception in the AMC *vs.* AOC conditions (resulting from bone conduction during vocalization) could have accounted for N1 amplitude differences, obscuring the interpretation of the findings. Therefore, the table only includes studies in which the same sounds were presented in the active and passive conditions: only the delay between an action and its feedback onset was manipulated.
[a] a) illusion (adaptation phase: fixed 200 ms delay; test phase: no delay); b) real-time (adaptation phase: no delay; test phase: no delay); c) delayed-time control (adaptation phase: fixed 200 ms delay; test phase: fixed same delay); d) deviant control (adaptation phase: fixed 400 ms delay; test phase: fixed 200 ms delay).

1.2. Prediction errors in response to the onset of self-generated sensory feedback

The button-press studies discussed in the previous section (Table 1) presented mixed findings regarding how the temporal delay between an action and its sensory feedback affects sensory attenuation. This can be attributed to how temporal predictability of sensory feedback is manipulated. In most of these studies, the delay of self-initiated stimulation was fixed (i.e., predictable). Baess et al. (2008) performed a study that examined the effects of prediction violations regarding a self-generated tone onset. In this study, delays between the action and the sound onset were fully unpredictable (i.e., the tone could be presented at any time between 500 and 1000 ms after the action). However, the authors did not investigate whether N1 suppression was modulated by the specific delays in which sensory feedback to an action occurred. Hence, it remains to be clarified how specific the efference copy is regarding the delay of expected sensory feedback. The remaining studies reported in Table 1 have used blocked designs to present delays whereby a delayed tone onset became predictable. These studies have not specifically addressed ERP responses to standard self-initiated sounds intermixed with deviants that violate expectations about sensory feedback. Furthermore, these studies have focused on the N1 as dependent variable and do not specify how manipulations of self-initiation and temporal predictability affect the time course of auditory processing beyond the N1.

The current study probed the effects of temporal predictability on the processing of self-generated *vs.* externally generated tones. When a stimulus is passively presented, the sound itself becomes less temporally predictable. In the current study, the active condition (auditory-motor condition) included trials in which the action effect was also not predictable in time to examine how temporal prediction errors (regarding the onset of sensory feedback) affect the processing of self-initiated tones. A previously established button-press paradigm (e.g., Knolle et al., 2012, 2013b) was adapted by comparing self-initiated with externally generated sounds, of which 30% were altered in timing. This allowed examining how specific temporal predictions for self-initiated tones are relative to externally generated tones. Specifically, we probed how delayed action-feedback modulates the evolution of ERP components associated with the comparison of actual and predicted feedback (P50, N1), the conscious detection of a self-generated sound (P2), error detection and update of an internal forward model (N2, P3). Tones were presented after a button press with no delay (standard/predictable condition – probability = 70%) or with a variable 50 ms, 100 ms, or 250 ms delay (deviant/unpredictable condition – probability = 30%). These specific delays were chosen based on evidence suggesting that the realization of an auditory efference copy cannot take longer than 150–170 ms (Whitford et al., 2011, but see Baess et al., 2008) and that a 300 ms delay is sufficient to abolish sensory attenuation (Tian and Poeppel, 2012). Moreover, the chosen delays allow testing the temporal integration window for successive events in auditory feedback processing (with an estimated duration of 200 ms; Clunies-Ross et al., 2015; Horváth et al., 2007; Näätänen et al., 2007; Wang et al., 2005).

Our first hypothesis concerned effects of self-initiation by comparing

ERP responses to self-*vs.* externally generated tones. Following prior studies (Knolle et al., 2013a, 2013b, 2012), we hypothesized that sounds presented with no delay after the button press (i.e., fully predictable onset time) should result in suppressed N1 and P2 responses, as predictions regarding onset time are fulfilled. Consistent with other studies, we expected to also observe sensory suppression to self-initiated tones in ERP components preceding the N1 (Baess et al., 2009).

Our second hypothesis concerned effects of action-feedback delay (temporal prediction errors) on sensory suppression. On the one hand, if the representation of the to-be-predicted sensation contains precise information about the onset of feedback, the suppression effect should be larger for specific predictions (i.e., non-delayed tone onset). As the sensory feedback deviates from the prediction of the internal forward model (by increasing the delay between the action and its sensory consequences), the error between predicted and perceived sensory feedback should increase, resulting in decreased N1 suppression. On the other hand, considering prior studies showing that sensory suppression relies on self-initiation more than on a precise temporal prediction about stimulus onset (Baess et al., 2008), sensory suppression should still be observed for self-generated tones presented with a 50 ms or 100 ms delay after the button press but far less with a 250 ms delay (that is, beyond the threshold of the temporal integration window of prediction and feedback [~200 ms] – e.g., Tian and Poeppel, 2015). This would be consistent with the idea that the internal forward model tolerates some degree of uncertainty in the timing of sensory feedback. However, beyond the temporal integration window (~200 ms), the auditory system should become more responsive to feedback that is not accurately predicted by the efference copy from the motor system (250 ms delay) and that is treated as unrelated to the action. Based on evidence suggesting that the P2 is more sensitive to the effects of temporal prediction (Sowman et al., 2012), we hypothesized an increase of the P2 amplitude the longer the delay between the action and its sensory consequences is.

Furthermore, if specific temporal predictions are generated but violated, the prediction error should be reflected in reduced sensory suppression after a deviant self-generated tone onset (N1, P2) as well as in enhanced ERP responses related to error detection, such as the N2 (Knolle et al., 2013b), reflecting the saliency of a self-generated deviant.

2. Method

2.1. Participants

Twenty-one college students (15 females) participated in the study ($M_{age} = 22.7$, $SD = 2.77$ years; age range 19–29 years). Participants were all right-handed (Oldfield, 1971), reported normal or corrected-to-normal visual acuity, normal hearing, no history of neurological illness, no personal history of psychiatric disorder or in first degree relatives, no present medication for a clinical condition that could affect EEG morphology or have neurological and/or cognitive functioning consequences. Before the EEG experiment, a brief clinical assessment ruled out the presence of psychopathological symptoms (Brief Symptom Inventory – Canavarro, 1999). None of the participants had total scores that indicated the presence of such symptoms.

All participants provided informed consent and were reimbursed for their time, either by course credits or a 10 € voucher. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local Ethics Committee of the University of Minho, Braga (Portugal).

2.2. Stimuli

A 680 Hz tone with 50 ms duration (Knolle et al., 2013b) was generated using Praat software (Boersma and Weenink, 2013). Tones were presented at ca. 70 dB sound pressure level via headphones (Sennheiser CX 300-II).

2.3. Procedure

In the EEG experiment each participant sat comfortably at a distance of about 1 m from a desktop computer monitor in a sound-attenuated and electrically shielded room. Each experimental session comprised three conditions (following Knolle et al., 2012, 2013b; 2013a): auditory-motor (AMC), auditory-only (AOC), and motor-only (MOC), described below and illustrated in Fig. 1. The three experimental blocks were preceded by a training block.

Motor Training: Participants practiced tapping every 2.4 s (sec) by imitating the interval played back by a metronome. Then, they were instructed to tap the sound sequence themselves. Visual feedback was displayed on the screen when their tap was too slow (>3 s) or too fast (<1.8 s). They had to produce the correct tap timing for 75% of trials (n = 50 trials) before they proceeded to the experimental session.

Auditory-Motor Condition (AMC): Participants produced finger taps approximately every 2.4 s (n = 280 standard/non-delayed and 120 deviant/delayed sounds), and each elicited a tone presented via headphones. In 30% of the trials the tone was presented with either a 50 ms (D50), 100 ms (D100), or 250 ms (D250) delay after the button press (n = 40 trials in each delayed condition). This percentage was chosen to increase the predictability of the standard (non-delayed – D0) sound and to ensure that the delayed sounds were perceived as deviants. The delay conditions were pseudorandomized with the constraint of a maximum of two deviant sounds of the same type presented before a new deviant type. The acoustic stimulation from the AMC was recorded on-line and used as the auditory sequence that was passively presented to participants in the AOC. This ensured that the sequence of sounds (and inter-sound intervals) was the same in the two critical conditions (AMC *vs.* AOC).

Auditory-Only Condition (AOC): Participants were instructed to listen attentively to the auditory stimuli played back at the exact same timing recorded from the AMC condition.

Motor-Only Condition (MOC): Participants performed self-paced finger taps every 2.4 s (n = 100 trials), but no sound was induced via the tap. The MOC served as a control condition that allowed monitoring of motor-related artifacts associated with the button press (AMC-MOC).

The AMC always preceded the AOC but the presentation of the MOC was randomized across participants. The presentation and timing of the stimuli was controlled by Presentation software (version 16.3; Neurobehavioral Systems, Inc.). A BioSemi tapping device was used to record the finger taps.

2.4. EEG data acquisition and analysis

EEG data were recorded using a 64-channel BioSemi Active Two system in continuous mode at a digitization rate of 512 Hz and stored on hard disk for later analysis.

Using Brain Vision Analyzer 2 (version 2.1.1.2516), EEG data were first re-referenced to the average reference and then band-pass filtered with a 0.1–90 Hz filter (zero phase shift Butterworth filters; low cutoff: 0.1 Hz, order 2; high cutoff: 30 Hz, order 2) for the analysis of N1 and later components. A more restrictive filter is usually recommended for the examination of early components preceding the N1 (Patterson et al., 2008). Following these suggestions, we applied a bandpass filter of 10–50 Hz (zero phase shift Butterworth filters; low cutoff: 10 Hz, order 8; high cutoff: 30 Hz, order 8) in a separate analysis probing P50 effects, following previous studies (Patterson et al., 2008; Rentzsch et al., 2008). However, we should note that, even though applying high-pass filters results in an improved signal-to-noise ratio, excessive filtering may also result in prominent signal distortions (Widmann et al., 2015). Hence, caution is warranted in data interpretation. The steps described next were the same for both types of analyses.

Individual ERP epochs (time-locked to sound onset), with -100 ms pre-stimulus baseline and 500 ms post-stimulus, were created for each condition. Tapping intervals shorter than 1.8 s or longer than 3.0 s were

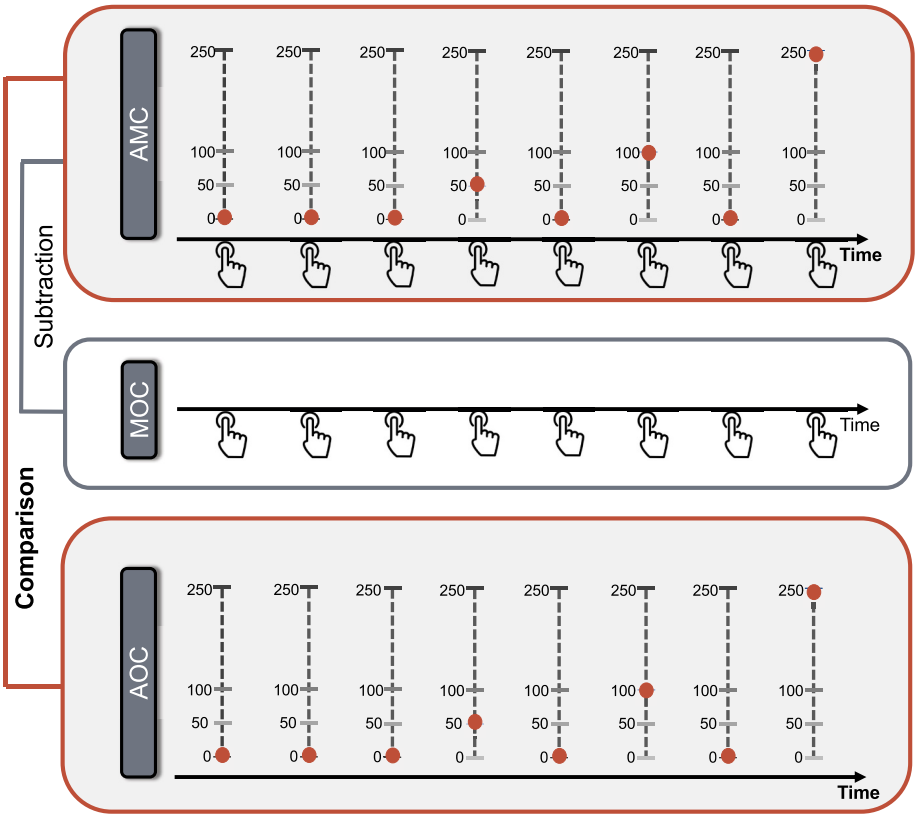

Fig. 1. Schematic illustration of the experimental design. Self-generated tones were presented immediately after the button press (D0: standard self-triggered tones – 70% of probability) or with a delay (D50, D100, D250: deviant self-generated tones – 30% of probability).

treated as errors, and the corresponding EEG recordings were excluded from further analyses (following Knolle et al., 2013a, 2013b). The EEG was baseline corrected using a -100 to 0 ms pre-stimulus interval. The vertical EOG was derived by subtracting the activity measured at an electrode positioned below the left eye from an electrode positioned above it. The horizontal EOG was derived by subtracting the activity measured at electrodes placed at the outer canthi of the eyes. Ocular artifacts were corrected using the algorithm of Gratton et al. (1983). EEG epochs with amplitudes exceeding±100 μV were rejected (mainly due to muscle artifacts, electrode drifting, and amplifier blocking). After artifact rejection, at least 75% of the segments per condition entered the analyses. Conditions did not differ in the number of non-rejected epochs

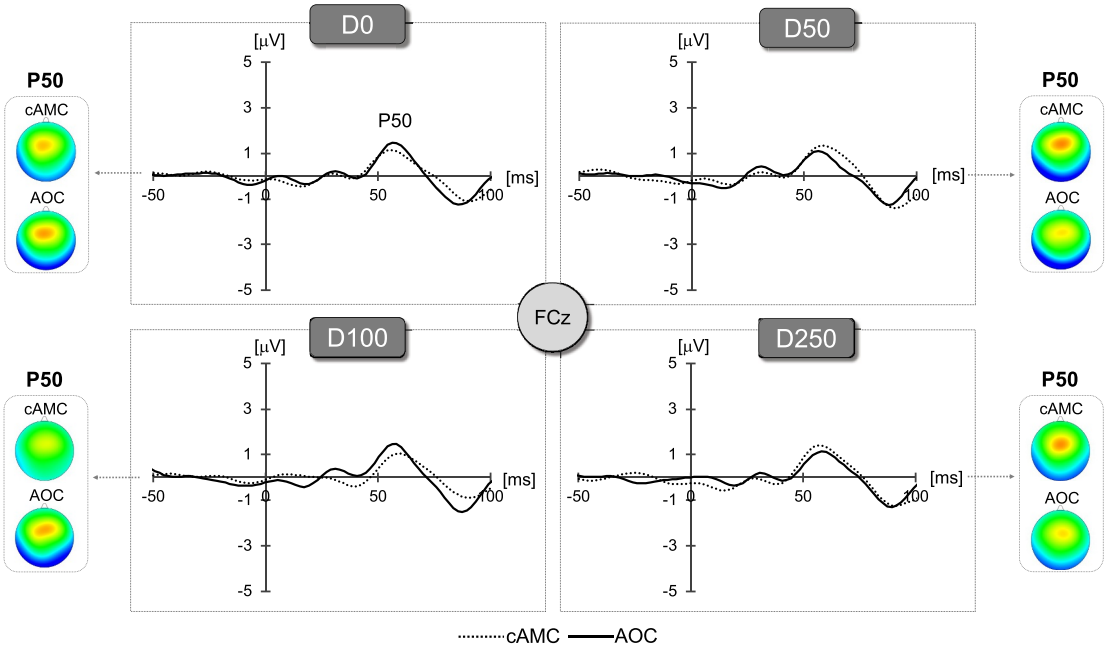

Fig. 2. The P50 response to self-generated and externally generated tones at each delay (D0, D50, D100, D250).

($p > .05$). Grand average waveforms were generated for each condition. Motor activity was subtracted from the AMC (corrected AMC [cAMC]) following prior studies (e.g., Baess et al., 2011; Elijah et al., 2016; Ford et al., 2014; Oestreich et al., 2016; SanMiguel et al., 2013). This allowed the comparison of sensory activity elicited in the AMC and AOC. All subsequent references to the AMC condition refer to the AMC condition after the MOC condition was subtracted (cAMC). To allow an equal distribution of the number of trials across conditions, only the non-rejected D0 trials preceding a given delay trial were included in the averages.

Visual inspection of the grand average waveforms (Figs. 2–4) revealed effects of the experimental manipulations in four components: a positive component peaking approximately 50 ms post-stimulus onset (P50), a negative component peaking at approximately 100 ms (N1), a positive component peaking at approximately 200 ms (P2) and a negative component peaking approximately at 280 ms (N2). Contrary to a prior study probing the effects of content prediction violations (unexpected pitch changes in the feedback) using a button-press task (Knolle et al., 2013b), the current grand average waveforms did not show a reliable P3 effect. Mean amplitudes were calculated for time windows lasting from 35 to 75 ms (P50), 80–120 ms (N1), 160–200 ms (P2) and 260–300 ms (N2), which broadly fit the latency windows used for the analysis of the same ERP components in prior studies (e.g., Knolle et al., 2013b; Mifsud and Whitford, 2017; Oestreich et al., 2016; Pinheiro et al., 2018; Rentzsch et al., 2008; Timm et al., 2016).

2.5. Statistical analysis

Following the inspection of the grand average waveforms (Figs. 2–4), and considering the typical frontocentral (N1 – Timm et al., 2013; N2 – Folstein and Van Petten, 2008) or central (P2 – Timm et al., 2013; N2 – Knolle et al., 2013b; Näätänen and Gaillard, 1983) topography of the expected ERP components, P50, N1, and N2 mean amplitudes were obtained from frontal (F3/Fz/F4) and frontocentral (FC3/FCz/FC4) electrodes, whereas P2 mean amplitude was obtained from central (C3/Cz/C4) electrodes.

Linear mixed effects models were built to fit the ERP amplitudes per participant and condition in each time-window of interest, using the lmer4 (Bates et al., 2015b) and lmerTest (Kuznetsova et al., 2016) packages in the R environment (R3.4.3. GUI 1.70). In contrast to the more traditional repeated-measures ANOVA analysis, mixed-effects modeling allows controlling for the variance associated with random factors such as random effects for participants in ERP amplitude (Baayen et al., 2008). Considering that mixed-effects modeling avoids spurious effects, can lead to more efficient estimates and has more power compared to traditional repeated-measures ANOVA (Bagiella et al., 2000; Boisgontier and Cheval, 2016; Jaeger, 2008), its use has been increasingly recommended in psychophysiology for the analysis of repeated-measures experiments (e.g., Bagiella et al., 2000). The default variance-covariance structure, i.e. the unstructured matrix, was used (Bates et al., 2015a). ERP signals from each participant were averaged across trials before they were included in the model due to the low signal-to-noise ratio of single-trial EEG data. The models included condition (cAMC *vs.* AOC), delay (D0, D50, D100, D250), and electrodes as fixed effects. Random effects terms included the subject-specific random intercepts and random slopes. The Satterthwaite approximation was applied to the REML-fitted models, which is preferred when evaluating significance in mixed-effects models (Luke, 2017).[3] The *anova* function from lmerTest (Kuznetsova et al., 2016) was used to provide *p*-values for each factor, calculated from the *F* statistic. Only significant effects ($p < .05$) are reported.

3. Results

3.1. Tapping performance

In the cAMC, the average tapping interval was 2364.85 ms ($SD = 214.49$ ms), with an overall correctness of 90.93% ($SD = 8.85\%$), not differing from normal distribution (Kolmogorov-Smirnov test, $p > .05$). In the MOC, the average tapping interval was 2514.47 ms ($SD = 478.44$ ms), with an overall correctness of 79.76% ($SD = 25.86\%$), not differing from normal distribution ($p > .05$).

3.2. ERP results

ERP waveforms showing ERP effects in the 4 cAMC and AOC conditions are presented in Figs. 2–4. Amplitude differences between the cAMC and AOC conditions (M±SEM) for each delay are presented in Fig. 5. Grand average waveforms contrasting MOC and AMC are shown in Supplementary Fig. 1.

3.2.1. Delayed action-feedback effects and the P50

The level of suppression was significantly affected by the type of delay (condition × delay interaction - F (3, 979) = 15.348, $p < .001$). Amplitude was less positive (i.e., suppressed) in response to self-generated (cAMC) than to externally generated (AOC) tones in the D0 ($\beta = 0.181$, SE = 0.057, t (979) = 3.144, $p = .002$; 95% CI: [0.068, 0.293]) and D100 delay conditions ($\beta = 0.127$, SE = 0.057, t (979) = 2.211, $p = .027$; 95% CI: [0.014, 0.239]; Figs. 2 and 5). In contrast, the P50 was increased (i.e., not suppressed) in response to self-initiated tones presented 50 ms (D50 – $\beta = -0.212$, SE = 0.057, t (979) = -3.684, $p < .001$; 95% CI: [-0.324, -0.099]) or 250 ms (D250 – $\beta = -0.181$, SE = 0.057, t (979) = -3.144, $p = .002$; 95% CI: [-0.293, -0.068]) after the button press.

3.2.2. Delayed action-feedback effects and the N1

As predicted, N1 suppression was affected by the type of delay (condition × delay interaction - F (3, 979) = 11.069, $p < .001$; Figs. 3 and 5). N1 was suppressed in response to self-initiated tones but only when they were presented up until 100 ms after the button press (D0 – $\beta = 0.974$, SE = 0.346, t (979) = 2.818, $p = .005$; 95% CI: [0.297, 1.651]; D50 – $\beta = 1.228$, SE = 0.346, t (979) = 3.555, $p < .001$; 95% CI: [0.551, 1.906]; D100 – $\beta = 1.968$, SE = 0.346, t (979) = 5.696, $p < .001$; 95% CI: [1.291, 2.645]). However, a reversed pattern was observed when tones were presented 250 ms after the button press: self-generated tones elicited an increased N1 amplitude compared to externally generated tones ($\beta = -0.974$, SE = 0.346, t (979) = -2.818, $p = .005$; 95% CI: [-1.651, -0.297]).

[3] Multilevel models incorporate in the same model the estimation of all parameters, analyzing them in parallel while relying on partial pooling. By more adequately modelling the within-group correlation structure of errors, they provide better estimates than classic multiple comparison corrections (Gelman et al., 2012; Gelman and Hill, 2007). The Satterthwaite approximation for degrees of freedom is not very susceptible to sample size and provides acceptable Type error I rates even when small samples are tested. Hence, besides the partial pooling underlying multilevel modelling, a non-sensitive *p*-value estimation was computed, both contributing to more reliable estimates in the current study (see also Supplementary Material).

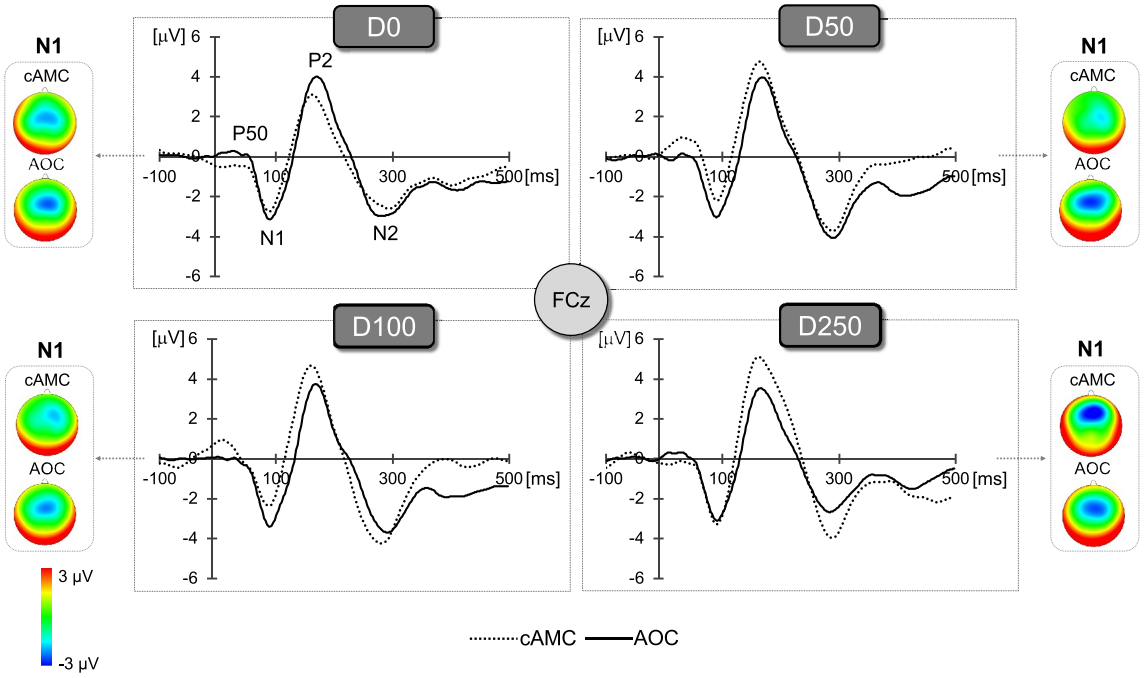

Fig. 3. Grand average waveforms (at FCz) contrasting ERP responses to self-generated and externally generated tones at each delay (D0, D50, D100, D250). Heat maps illustrating the topographical distribution of the N1 effect are shown for each delay.

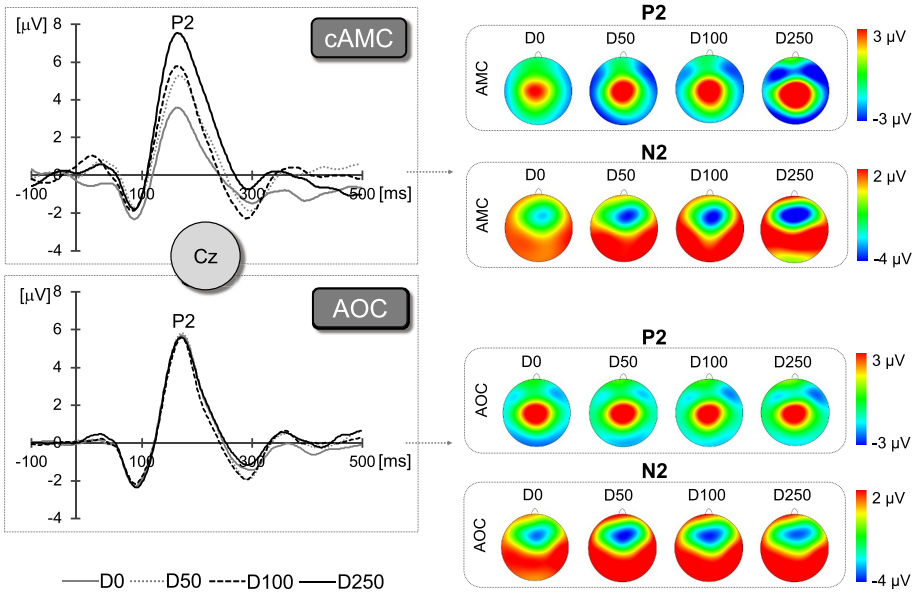

Fig. 4. The P2 and N2 responses to self-generated and externally generated tones in each delay (D0, D50, D100, D250). Heat maps illustrating the topographical distribution of the P2 and N2 effects are shown for each delay.

3.2.3. *Delayed action-feedback effects and the P2*

The effect of condition interacted with delay (F (3, 476) = 18.340, $p < .001$; Figs. 4 and 5). P2 amplitude in response to self-generated tones increased as a function of delay, but not in response to externally generated tones (in the AOC, the effect of delay was not significant - F (3, 228) = 1.757, $p = .156$). A gradual increase in P2 amplitude was observed for self-initiated compared to externally triggered tones as a function of the action-tone delay (D50 – β = 1.678, SE = 0.769, t (476) = 2.182, $p = .030$; 95% CI: [0.171, 3.184]; D100 – β = 2.021, SE = 0.769, t (476) = 2.629, $p = .009$; 95% CI: [0.514, 3.528]; D250 – β = 5.541, SE = 0.769, t (476) = 7.208, $p < .001$; 95% CI: [4.034, 7.048]).

Specifically, for sounds that were elicited by a button press, the P2 was increased in the D250 compared to D0 delay (β = 5.541, SE = 0.769, t (476) = 7.208, $p < .001$; 95% CI: [4.034, 7.048]), D50 delay (β = 3.864, SE = 0.769, t (476) = 5.025, $p < .001$; 95% CI: [2.357, 5.370]) and D100 delay (β = 3.520, SE = 0.769, t (476) = 4.579, $p < .001$; 95% CI: [2.013, 5.027]). P2 was also more positive in the D100 compared to D0 delay (β = 2.021, SE = 0.769, t (476) = 2.629, $p = .009$; 95% CI: [0.514, 3.528]) and in the D50 compared to D0 delay (β = 1.678, SE = 0.769, t (476) = 2.182, $p = .030$; 95% CI: [0.171, 3.184]).

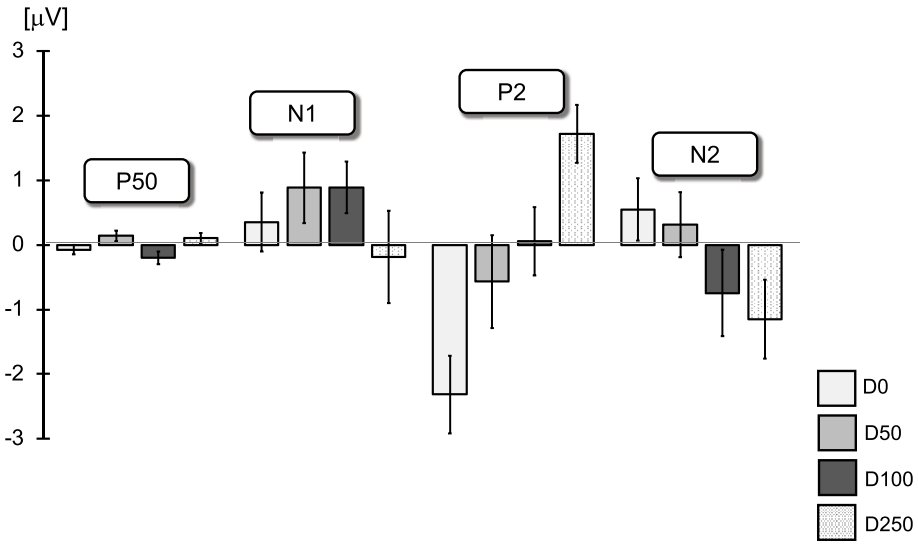

Fig. 5. ERP amplitude differences between the cAMC and AOC conditions (for the P50 and P2, negative values represent a less positive amplitude for the cAMC compared to AOC, whereas positive values represent a more positive amplitude for the cAMC compared to AOC; for the N1 and N2, negative values represent more negative amplitude for the cAMC compared to AOC, whereas positive values represent less negative amplitude for the cAMC compared to AOC) for each delay. Bars represent mean amplitudes over FCz in the case of P50, N1 and N2, and over Cz in the case of P2. Standard error (SE) of the means is shown in error bars.

3.2.4. *Delayed action-feedback effects and the N2*

Condition interacted with delay in modulating N2 amplitude (F (3, 979) = 6.692, $p < .001$). Self-generated and externally generated sounds were processed differently in the D250 delay condition only (β = -1.512, SE = 0.363, t (979) = -4.165, $p < .001$; 95% CI: [-2.223, -0.800]; Figs. 4 and 5): amplitude was more negative for self-generated compared to externally generated tones. The N2 elicited by tones presented 250 ms after the button press was significantly more negative than in the other delay conditions (D0 – β = -1.512, SE = 0.363, t (979) = -4.165, $p < .001$; 95% CI: [-2.223, -0.800]; D50 – β = -1.238, SE = 0.363, t (979) = -3.410, $p < .001$; 95% CI: [-1.950, -0.526]; D100 – β = -1.100, SE = 0.363, t (979) = -3.029, $p = .003$; 95% CI: [-1.811, -0.388]).

4. Discussion

The current study aimed to specify the flexibility of the efference copy regarding the onset of sensory feedback to self-generated action. Further, it examined how manipulations of the onset delay, leading to prediction errors, affect distinct stages of auditory processing. We replicated the sensory suppression effect reported in prior studies (Baess et al., 2011; Knolle et al., 2013b, 2013a; 2012; Timm et al., 2016): self-generated tones elicited a significantly smaller N1 response than externally generated tones. We extended this evidence by showing suppression effects arising before the N1, starting within 50 ms post-feedback onset (P50). In the P2 latency range, sound processing became sensitive to delay differences between an action and its sensory feedback: P2 amplitude increased proportionally to the size of the delay of self-generated tones.

Our findings also indicate that temporal predictions regarding the onset of a self-generated tone are somewhat flexible: self-generated deviant sounds presented 100 ms after an action elicited a significant P50 and N1 sensory suppression. Contrary, self-generated sounds presented 250 ms after an action did not lead to N1 and P2 suppression, but elicited a N2 response associated with the conscious detection of an error due to delayed sensory feedback (Fig. 6). In the following we discuss these findings chronologically.

4.1. *Predicted and perceived sensory feedback: P50 and N1 effects*

Sensory suppression started before the N1 as indicated by a suppressed P50 response to self-generated tones presented concomitantly with the button press. Suppression effects occurring approximately at

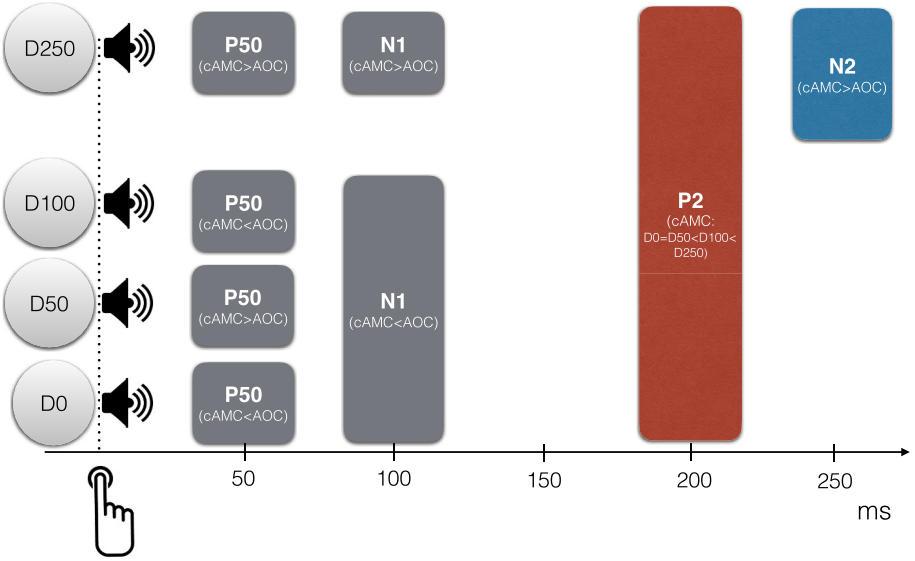

Fig. 6. Schematic illustration of the effects of self-initiation and action-feedback delays in the current study. The effects of self-initiation were more prominent within the first 150 ms post-tone onset (grey: P50, N1), whereas the P2 was particularly sensitive to the timing of sensory feedback for action (red). Tones presented with the longest delay (250 ms) after the action (blue) elicited an additional negativity (N2): N2 was more negative for self-generated compared to externally generated tones. D0 = 0 ms delay; D50 = 50 ms delay; D100 = 100 ms delay; D250 = 250 ms delay. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

50 ms after a tone onset have been reported in the context of gating paradigms (Boutros and Belger, 1999) as well as in studies that manipulated the temporal predictability of sounds (Schwartze et al., 2013; White and Yee, 2006). The latter studies confirmed a decreased P50 for more predictable than less predictable sounds (Schwartze et al., 2013; White and Yee, 2006). Further, a prior study using a button press paradigm (with no manipulations of temporal predictability) reported an attenuation already in the middle latency range (Pa, Nb), occurring at 27–33 ms and 57–63 ms respectively (Baess et al., 2009). The current findings suggest that sensory feedback to a self-action starts within 50 ms after a sound onset, a time-window spanning brainstem and thalamic levels of sound processing (e.g., Grimm and Escera, 2012). These earlier effects align with a hierarchical predictive coding view that postulates an interplay between stimulus-driven (forward) and model-driven (backward) connections from the earliest stages of feedback processing (e.g., Grimm and Escera, 2012).

Contrary to our prediction, the P50 was not suppressed, but enhanced, when sounds were presented 50 ms or 250 ms after the button press. The lack of P50 suppression for the two delay conditions confirms that the auditory system is sensitive to the temporal predictability of a sound from early processing stages of auditory information in the human brain onwards (Baess et al., 2009; Schwartze et al., 2013; White and Yee, 2006). However, the enhanced P50 response to self-initiated tones presented 50 ms after the button press seems counterintuitive considering that these same tones were associated with a suppressed N1 response. Future studies should clarify why attenuation in the processing of delayed self-initiated sounds was observed only when tones are presented 100 ms after the button press. Considering that restrictive high-pass filters may introduce prominent signal distortions not only on the temporal dynamics but also on component topographies and morphology (e.g., reduced peak amplitude; see Widmann et al., 2015), the P50 findings (suppressed P50 when tones are presented 100 ms after the button press [cAMC < AOC] and enhanced P50 when tones are presented 50 or 250 ms after the button press [cAMC > AOC]) should thus be interpreted with caution.

The N1 findings replicate the classical sensory suppression effect: self-generated tones elicited a smaller N1 response than externally generated tones. This finding corroborates the notion that the efference copy of motor commands contains timing-specific information (i.e., *when* sensory feedback to self-action should occur). This was previously seen in suppression effects in unisensory processing (Schwartze et al., 2013). Critically, N1 suppression was still observed when self-generated sounds were presented with a delay of 50 or 100 ms after the button press (D50 and D100). However, the N1 suppression effect was eliminated with a delay of 250 ms after the button press.

Contrary to the hypothesis that the specificity of temporal prediction would be inversely related to the amount of neural suppression, we did not observe a stronger suppression effect for those predictions that were temporally more precise (i.e., standard self-generated tones presented with no delay compared to deviant tones presented with a 50 ms or a 100 ms delay). The suppression effect observed in response to sounds presented with a 50 ms or 100 ms delay after an action speaks against the hypothesis that the suppression effect primarily reflects the coinciding button press rather than a prediction effect of when a sound should occur (Elijah et al., 2016; Horváth et al., 2012; Oestreich et al., 2016). This observation concurs with some degree of flexibility when sensory suppression occurs in auditory regions, where the sensory N1 subcomponents are thought to be generated (Woods, 1995). The efference copy of motor commands is thought to be temporally specific, resulting in cancellation of neural responses to auditory feedback that is temporally more predictable due to one's own action. Notwithstanding, the internal forward model may tolerate some degree of uncertainty regarding stimulus onset characteristics (Baess et al., 2008; Lange, 2011). This may occur because inflexibly precise predictions may be disadvantageous for extracorporeal sensory feedback to actions (e.g., tones) due to the high variance or noise in their sensory outcome (Gentsch et al., 2012) and to avoid unnecessary motor correction (Tian and Poeppel, 2015).

Studies using speech production *vs.* passive listening showed that sensory suppression occurs when sensory feedback is presented within 200 ms of a vocalization (Behroozmand and Larson, 2011), which agrees with the current findings. The 200 ms window after the button press seems to be critical for action-sound integration (corresponding to the temporal window of integration in auditory perception – e.g., Horváth et al., 2007; Näätänen et al., 2007; Wang et al., 2005). The temporal distance between prediction and feedback was found to modulate the N1 amplitude. Consistent with the current findings, Okayasu and collaborators (Okayasu et al., 2019) observed that bone-conducted ultrasounds are bound within a temporal integration window with a stimulus onset asynchrony (SOA) of 100–150 ms, but are perceived as separate events if the SOA has a 200 or 350 ms duration. Similar effects of reduced attenuation to action-feedback delays longer than 200 ms were observed in tactile processing (Blakemore et al., 1999) and auditory processing using speech generation paradigms (Tian and Poeppel, 2015).

4.2. Detection of a self-generated sound: P2 effects

N1 and P2 suppression effects are functionally dissociable (e.g., N1 suppression, but not P2 suppression, is affected in cerebellar patients – Knolle et al., 2012, 2013a). Even though suppression effects observed in the P50 and in the N1 were not further modulated by specific differences in the action-feedback delay (i.e., the difference observed was between any tone presented within 100 ms or after 100 ms following an action), the P2 findings revealed a graded response. Specifically, P2 amplitude increased with longer delays between the button press and the respective sensory feedback. Imposing a 250 ms delay may have caused the self-generated feedback to be processed in the same way as input generated by an external source (i.e., decreasing the sense of agency). Importantly, even though the timing of externally generated tones was the same as for self-generated tones, the effect of the delay was not significant in the case of the former (condition by delay interaction), which suggests that the auditory system is more responsive to perturbations in the timing of sensory feedback to action compared to passive listening. The current findings agree with an enhanced P2 response when the timing of auditory feedback was unpredictable (Behroozmand et al., 2016; Chen et al., 2012).

In studies using the button press paradigm, P2 reduction has been linked to conscious detection of a self-initiated sound (Knolle et al., 2013b, 2013a, 2012) and to be more directly related to behavioral indices of agency perception (Timm et al., 2016). For example, in previous studies the perceived saliency of self-generated tactile stimulation increased with an increasing delay between an action and sensory feedback (Blakemore et al., 1999). The temporal proximity between the action (button press) and the sound is a relevant cue for agency attribution. Increased P2 for self-initiated sounds presented 250 ms after the button press may thus reflect decreased agency attribution or the conscious detection that these sounds were not self-generated (Knolle et al., 2013b; Sowman et al., 2012). The P2 increase may serve an adaptive function leading to updates of the internal model and action correction (e.g., Behroozmand et al., 2011).

The enhanced delay sensitivity of the P2 suggests that, whereas the N1 may monitor the suppression effect of the efference copy, the P2 reflects the suppression effect of self-initiation that is contingent on the temporal proximity between an action and sensory feedback (Sowman et al., 2012). It should be noted that the delay effects tested here and in previous studies may be modality- and stimulus-specific (Mifsud and Whitford, 2017).

4.3. Voluntary error detection: N2 effects

Next to the reduced suppression effect observed in the P50, N1 and P2 components, sensory feedback at a 250 ms delay elicited a larger N2

to self-generated than externally generated tones. This N2 enhancement has been associated with voluntary deviance detection (Knolle et al., 2013b). Specifically, the N2 has been related to the conscious detection of an infrequent variation of stimulus properties (Horváth et al., 2008; Ritter et al., 1992) and may lead to more efficient processing given that more processing resources are activated (Knolle et al., 2013b). In a previous study using a button-press paradigm, the N2 was more enhanced for self-generated than externally generated tones, which was interpreted as evidence that unexpected sensory feedback is salient, leads to conscious error awareness, and calls for enhanced resources allocation (Knolle et al., 2013b). This indicates that, in this context, the perturbation of a temporal prediction becomes salient only when tones are presented with a 250 ms delay. It should be noted that self-generated tones with such a delay did not elicit attentional orienting (as previously observed for self-generated tones with infrequent pitch changes – Knolle et al., 2013b), which would lead to a P3a response. This represents further support for the notion that content (*what*) and temporal (*when*) predictions are subserved by different neurofunctional mechanisms (Hsu et al., 2013).

4.4. Implications for current accounts of error processing in the brain

Generally, a prediction error is defined as the difference between a predicted and the incoming sensory signal. This difference reflects the level of surprise arising from the comparison of expected and actual outcomes (Friston, 2010). Different types of prediction errors have been proposed, namely perceptual, cognitive, and motivational prediction errors (Den Ouden et al., 2012). The type of prediction error addressed in the current study fits the category of perceptual prediction errors.

According to the free energy principle (e.g., Friston, 2009), the amount of prediction error is proportional to the amount of surprise in response to sensory input that cannot be explained by an internal forward model. Only the difference between prediction and incoming input is thought to be transmitted to higher cortical areas (Wacongne et al., 2011). ERP components with a distinct latency have been found to respond to prediction violations in various tasks, including the MMN (peaking approximately at 100–200 ms post-stimulus onset; e.g., Wacongne et al., 2011), P300 (peaking approximately at 300 ms post-stimulus onset; e.g., Van Petten and Luka, 2012; Wacongne et al., 2011), N400 (peaking approximately at 400 ms post-stimulus onset; Van Petten and Luka, 2012), or P600 (peaking approximately at 600 ms post-stimulus onset; e.g., Van Petten and Luka, 2012). These ERP components might to various degrees index the difference between the expected and actual incoming sensory signal at multiple processing stages (Kiebel et al., 2008; Wacongne et al., 2011), consistent with the observation that the coding of prediction errors is ubiquitous in the brain (Den Ouden et al., 2012).

The current findings agree with the notion that the detection of prediction errors appears to be organised in several stages, and that the prediction error that arises at a given stage may serve as the input to the next one (e.g., Wacongne et al., 2011). Additionally, they suggest that different action-feedback delays may vary the degree of a prediction error (or its distinct precision), with a differential impact upon evoked potentials indexing distinct stages of error processing.

4.5. Limitations and future directions

In the current study, motor activity was subtracted from the AMC (cAMC). This subtraction approach has been consistently used in prior studies (e.g., Baess et al., 2011; Elijah et al., 2016; Ford et al., 2014; Oestreich et al., 2016; SanMiguel et al., 2013) and allows for comparability with the current findings. Even though no significant differences were observed in the baseline activity before sounds that were self-initiated (after motor correction – cAMC) *vs.* externally triggered (see Supplementary Material), we cannot fully rule out any contribution of motor activity to sound processing in the cAMC condition. Previous studies found that brain activity before action onset was modulated by the expectation of self-generated sensory consequences (e.g., Reznik et al., 2018; Vercillo et al., 2018). Specifically, the Readiness Potential preceding a button press was increased (i.e., more negative) when the action produced a sensory effect (AMC) compared to actions with no sensory consequences (MOC) (Reznik et al., 2018; Vercillo et al., 2018). These findings support the relevance of examining brain activity preceding stimulus onset for a full understanding of the forward model function (e.g., Reznik et al., 2018). Future studies should specify the effects of delays on brain activity preceding sound onset (e.g., see Reznik et al., 2018).

5. Conclusions

The current ERP findings reveal that specific predictions concerning the temporal onset of self-generated tones are generated, despite tolerating temporal uncertainty of up to 100 ms. The P50 was suppressed in response to self-initiated tones, but only when they are presented immediately or 100 ms after the button press; a P50 enhancement effect was observed when tones are presented 50 or 250 ms after the button press. The comparison of predicted and actual sensory feedback (reflected in N1 amplitude modulations) was constrained by a temporal integration window, with a duration of approximately 100 ms: non-delayed and delayed (50 ms, 100 ms) self-generated tones were processed similarly. In the P2 latency range, sound processing becomes more sensitive to delay differences between an action and its sensory feedback, which may impact the conscious detection of altered action-related sensory feedback. Sensory feedback presented with a 250 ms delay (occurring after a temporal integration window) elicited further ERP responses associated with error detection, i.e. an enhanced N2, which may lead to conscious error awareness.

These findings suggest functionally dissociable steps in how temporal prediction modulates the internal forward model. Further, they indicate that different types of temporal prediction errors are treated differently by the auditory system. Together, they shed light on the cognitive mechanisms mediating temporal prediction in auditory feedback processing, with implications for our understanding of the role of brain regions involved in forward prediction and rapid transmission of event-based representation of temporal structure, such as the cerebellum (Schwartze and Kotz, 2013).

CRediT authorship contribution statement

Ana P. Pinheiro: Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Software, Supervision, Visualization, Writing - original draft, Writing - review & editing. **Michael Schwartze:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Writing - review & editing. **Francisco Gutierrez:** Data curation, Formal analysis, Writing - review & editing. **Sonja A. Kotz:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Acknowledgments

This work was supported by the Portuguese Science Foundation (*Fundação para a Ciência e a Tecnologia* [FCT]; grant numbers IF/00334/2012, PTDC/MHC-PCN/0101/2014 awarded to APP, MS [PTDC/MHC-PCN/0101/2014], and SAK [PTDC/MHC-PCN/0101/2014]) and BIAL Foundation (BIAL 238/16). The Authors gratefully acknowledge all the participants who collaborated in the study, and particularly Dr. Franziska Knolle for feedback on the stimulus generation and Dr. Magda Roberto for statistical consultancy.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2019.107200.

References

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59, 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Baess, P., Horvath, J., Jacobsen, T., Schroger, E., 2011. Selective suppression of self-initiated sounds in an auditory stream: an ERP study. Psychophysiology 48, 1276–1283. https://doi.org/10.1111/j.1469-8986.2011.01196.x.

Baess, P., Jacobsen, T., Schroger, E., 2008. Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. Int. J. Psychophysiol. 70, 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005.

Baess, P., Widmann, A., Roye, A., Schröger, E., Jacobsen, T., 2009. Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. Eur. J. Neurosci. 29, 1514–1521. https://doi.org/10.1111/j.1460-9568.2009.06683.x.

Bagiella, E., Sloan, R.P., Heitjan, D.F., 2000. Mixed-effects models in psychophysiology. Psychophysiology 37, 13–20. https://doi.org/10.1017/S0048577200980648.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. lme4: linear mixed-effects models using Eigen and S4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Behroozmand, R., Karvelis, L., Liu, H., Larson, C.R., 2009. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clin. Neurophysiol. 120, 1303–1312. https://doi.org/10.1016/j.clinph.2009.04.022.

Behroozmand, R., Larson, C.R., 2011. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci. 12, 54. https://doi.org/10.1186/1471-2202-12-54.

Behroozmand, R., Liu, H., Larson, C.R., 2011. Time-dependent neural processing of auditory feedback during voice pitch error detection. J. Cogn. Neurosci. 23, 1205–1217. https://doi.org/10.1162/jocn.2010.21447.

Behroozmand, R., Sangtian, S., Korzyukov, O., Larson, C.R., 2016. A temporal predictive code for voice motor control: evidence from ERP and behavioral responses to pitch-shifted auditory feedback. Brain Res. 1636, 1–12. https://doi.org/10.1016/j.brainres.2016.01.040.

Blakemore, S.-J., Frith, C.D., Wolpert, D.M., 1999. Spatio-temporal prediction modulates the perception of self-produced stimuli. J. Cogn. Neurosci. 11, 551–559. https://doi.org/10.1162/089892999563607.

Boersma, P., Weenink, D., 2013. Praat: Doing Phonetics by Computer [Computer Program], Version 5.3.53.

Boisgontier, M.P., Cheval, B., 2016. The anova to mixed model transition. Neurosci. Biobehav. Rev. 68, 1004–1005. https://doi.org/10.1016/j.neubiorev.2016.05.034.

Boutros, N.N., Belger, A., 1999. Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. Biol. Psychiatry 45, 917–922. https://doi.org/10.1016/S0006-3223(98)00253-4.

Canavarro, M.C., 1999. Inventário de Sintomas psicopatológicos [brief symptoms inventory]: BSI. In: Simões, M.R., Gonçalves, M., LSA (Eds.), Testes e Provas Psicológicas Em Portugal [Psychological Tests in Portugal]-Vol. II. SHO/APPORT, Braga, pp. 87–109.

Chen, Z., Chen, X., Liu, P., Huang, D., Liu, H., 2012. Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. BMC Neurosci. 13, 55. https://doi.org/10.1186/1471-2202-13-55.

Clunies-Ross, K.L., Brydges, C.R., Nguyen, A.T., Fox, A.M., 2015. Hemispheric asymmetries in auditory temporal integration: a study of event-related potentials. Neuropsychologia 68, 201–208. https://doi.org/10.1016/j.neuropsychologia.2015.01.018.

Crowley, K.E., Colrain, I.M., 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin. Neurophysiol. 115, 732–744. https://doi.org/10.1016/j.clinph.2003.11.021.

Den Ouden, H.E.M., Kok, P., de Lange, F.P., 2012. How prediction errors shape perception, attention, and motivation. Front. Psychol. 3, 548. https://doi.org/10.3389/fpsyg.2012.00548.

Elijah, R.B., Le Pelley, M.E., Whitford, T.J., 2016. Modifying temporal expectations: changing cortical responsivity to delayed self-initiated sensations with training. Biol. Psychol. 120, 88–95. https://doi.org/10.1016/j.biopsycho.2016.09.001.

Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology 45, 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x.

Ford, J.M., Mathalon, D.H., Heinks, T., Kalba, S., Faustman, W.O., Roth, W.T., 2001. Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. Am. J. Psychiatry 158, 2069–2071. https://doi.org/10.1176/appi.ajp.158.12.2069.

Ford, J.M., Mathalon, D.H., Kalba, S., Whitfield, S., Faustman, W.O., Roth, W.T., 2001. Cortical responsiveness during talking and listening in schizophrenia: an event-related brain potential study. Biol. Psychiatry 50, 540–549. https://doi.org/10.1016/S0006-3223(01)01166-0.

Ford, J.M., Palzes, V.A., Roach, B.J., Mathalon, D.H., 2014. Did i do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. Schizophr. Bull. 40, 804–812. https://doi.org/10.1093/schbul/sbt072.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138. https://doi.org/10.1038/nrn2787.

Friston, K., 2009. The free-energy principle: a rough guide to the brain? Trends Cogn. Sci. 13, 293–301. https://doi.org/10.1016/j.tics.2009.04.005.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 815–836. https://doi.org/10.1098/rstb.2005.1622.

Gelman, A., Hill, J., Yajima, M., 2012. Why we (usually) don't have to worry about multiple comparisons. J. Res. Educ. Eff. 5, 189–211. https://doi.org/10.1080/19345747.2011.618213.

Gelman, A., Hill, J.L., 2007. Data Analysis Using Regression and Multilevel/hierarchical Models. Cambridge University Press, Cambridge, UK. https://doi.org/10.2277/0521867061.

Gentsch, A., Kathmann, N., Schütz-Bosbach, S., 2012. Reliability of sensory predictions determines the experience of self-agency. Behav. Brain Res. 228, 415–422. https://doi.org/10.1016/j.bbr.2011.12.029.

Godey, B., Schwartz, D., De Graaf, J.B., Chauvel, P., Liégeois-Chauvel, C., 2001. Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clin. Neurophysiol. 112, 1850–1859. https://doi.org/10.1016/S1388-2457(01)00636-8.

Gorina-Careta, N., Zarnowiec, K., Costa-Faidella, J., Escera, C., 2016. Timing predictability enhances regularity encoding in the human subcortical auditory pathway. Sci. Rep. 6, 37405. https://doi.org/10.1038/srep37405.

Gratton, G., Coles, M.G.H., Donchin, E., 1983. A new method for off-line removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. 55, 468–484. https://doi.org/10.1016/0013-4694(83)90135-9.

Grimm, S., Escera, C., 2012. Auditory deviance detection revisited: evidence for a hierarchical novelty system. Int. J. Psychophysiol. 85, 88–92. https://doi.org/10.1016/j.ijpsycho.2011.05.012.

Holst, E., Mittelstaedt, H., 1950. Das Reafferenzprinzip, Die Naturwissenschaften. https://doi.org/10.1007/BF00622503.

Horváth, J., 2015. Action-related auditory ERP attenuation: paradigms and hypotheses. Brain Res. 1626, 54–65. https://doi.org/10.1016/j.brainres.2015.03.038.

Horváth, J., Czigler, I., Winkler, I., Teder-Sälejärvi, W.A., 2007. The temporal window of integration in elderly and young adults. Neurobiol. Aging 28, 964–975. https://doi.org/10.1016/j.neurobiolaging.2006.05.002.

Horváth, J., Maess, B., Baess, P., Tóth, A., 2012. Action–sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. J. Cogn. Neurosci. 24, 1919–1931. https://doi.org/10.1162/jocn_a_00215.

Horváth, J., Roeber, U., Bendixen, A., Schröger, E., 2008. Specific or general? The nature of attention set changes triggered by distracting auditory events. Brain Res. 1229, 193–203. https://doi.org/10.1016/j.brainres.2008.06.096.

Hsu, Y.F., Hämäläinen, J.A., Waszak, F., 2013. Temporal expectation and spectral expectation operate in distinct fashion on neuronal populations. Neuropsychologia 51, 2548–2555. https://doi.org/10.1016/j.neuropsychologia.2013.09.018.

Jaeger, T.F., 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. J. Mem. Lang. 59, 434–446. https://doi.org/10.1016/j.jml.2007.11.007.

Kiebel, S.J., Daunizeau, J., Friston, K.J., 2008. A hierarchy of time-scales and the brain. PLoS Comput. Biol. 4, e1000209. https://doi.org/10.1371/journal.pcbi.1000209.

Knolle, F., Schröger, E., Baess, P., Kotz, S.A., 2012. The cerebellum generates motor-to-auditory Predictions: ERP lesion evidence. J. Cogn. Neurosci. 24, 698–706. https://doi.org/10.1162/jocn_a_00167.

Knolle, F., Schröger, E., Kotz, S.A., 2013. Cerebellar contribution to the prediction of self-initiated sounds. Cortex 49, 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012.

Knolle, F., Schröger, E., Kotz, S.A., 2013. Prediction errors in self- and externally-generated deviants. Biol. Psychol. 92, 410–416. https://doi.org/10.1016/j.biopsycho.2012.11.017.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2016. lmerTest: tests in linear mixed effects models. R Packag. version 2, 0–33.

Lange, K., 2011. The reduced N1 to self-generated tones: an effect of temporal predictability? Psychophysiology 48, 1088–1095. https://doi.org/10.1111/j.1469-8986.2010.01174.x.

Leung, S., Recasens, M., Grimm, S., Escera, C., 2013. Electrophysiological index of acoustic temporal regularity violation in the middle latency range. Clin. Neurophysiol. 124, 2397–2405. https://doi.org/10.1016/j.clinph.2013.06.001.

Luke, S.G., 2017. Evaluating significance in linear mixed-effects models in R. Behav. Res. Methods 49, 1494–1502. https://doi.org/10.3758/s13428-016-0809-y.

Mates, J., Müller, U., Radil, T., Pöppel, E., 1994. Temporal integration in sensorimotor synchronization. J. Cogn. Neurosci. 6, 332–340. https://doi.org/10.1162/jocn.1994.6.4.332.

Mifsud, N.G., Whitford, T.J., 2017. Sensory attenuation of self-initiated sounds maps onto habitual associations between motor action and sound. Neuropsychologia 103, 38–43. https://doi.org/10.1016/j.neuropsychologia.2017.07.019.

Näätänen, R., Gaillard, A.W.K., 1983. The orienting reflex and the N2 deflection of the event-related potential (ERP). Adv. Psychol. 10, 119–141. https://doi.org/10.1016/S0166-4115(08)62036-1.

Näätänen, R., Michie, P.T., 1979. Early selective-attention effects on the evoked potential: a critical review and reinterpretation. Biol. Psychol. 8, 81–136. https://doi.org/10.1016/0301-0511(79)90053-X.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118, 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026.

Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure.

Psychophysiology 24, 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x.
Oestreich, L.K.L., Mifsud, N.G., Ford, J.M., Roach, B.J., Mathalon, D.H., Whitford, T.J., 2016. Cortical suppression to delayed self-initiated auditory stimuli in schizotypy. Clin. EEG Neurosci. 47, 3–10. https://doi.org/10.1177/1550059415581708.
Okayasu, T., Nishimura, T., Uratani, Y., Yamashita, A., Nakagawa, S., Yamanaka, T., Hosoi, H., Kitahara, T., 2019. Temporal window of integration estimated by omission in bone-conducted ultrasound. Neurosci. Lett. 696, 1–6. https://doi.org/10.1016/j.neulet.2018.11.035.
Patterson, J.V., Hetrick, W.P., Boutros, N.N., Jin, Y., Sandman, C., Stern, H., Potkin, S., Bunney, W.E., 2008. P50 sensory gating ratios in schizophrenics and controls: a review and data analysis. Psychiatry Res. 158, 226–247. https://doi.org/10.1016/j.psychres.2007.02.009.
Perez, V.B., Ford, J.M., Roach, B.J., Loewy, R.L., Stuart, B.K., Vinogradov, S., Mathalon, D.H., 2012. Auditory cortex responsiveness during talking and listening: early illness schizophrenia and patients at clinical high-risk for psychosis. Schizophr. Bull. 38, 1216–1224. https://doi.org/10.1093/schbul/sbr124.
Pinheiro, A.P., Schwartze, M., Kotz, S.A., 2018. Voice-selective prediction alterations in nonclinical voice hearers. Sci. Rep. 8, 14717. https://doi.org/10.1038/s41598-018-32614-9.
Rentzsch, J., Jockers-Scherübl, M.C., Boutros, N.N., Gallinat, J., 2008. Test-retest reliability of P50, N100 and P200 auditory sensory gating in healthy subjects. Int. J. Psychophysiol. 67, 81–90. https://doi.org/10.1016/j.ijpsycho.2007.10.006.
Reznik, D., Simon, S., Mukamel, R., 2018. Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. Neuropsychologia 119, 302–307. https://doi.org/10.1016/j.neuropsychologia.2018.08.028.
Ritter, W., Paavilainen, P., Lavikainen, J., Reinikainen, K., Alho, K., Sams, M., Näätänen, R., 1992. Event-related potentials to repetition and change of auditory stimuli. Electroencephalogr. Clin. Neurophysiol. 83, 306–321. https://doi.org/10.1016/0013-4694(92)90090-5.
SanMiguel, I., Todd, J., Schröger, E., 2013. Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. Psychophysiology 50, 334–343. https://doi.org/10.1111/psyp.12024.
Schafer, E.W.P., Marcus, M.M., 1973. Self-stimulation alters human sensory brain responses. Science 181, 175–177. https://doi.org/10.1126/science.181.4095.175.
Schroger, E., Kotz, S.A., SanMiguel, I., 2015. Bridging prediction and attention in current research on perception and action. Brain Res. 1626, 1–13. https://doi.org/10.1016/j.brainres.2015.08.037.
Schwartze, M., Farrugia, N., Kotz, S.A., 2013. Dissociation of formal and temporal predictability in early auditory evoked potentials. Neuropsychologia 51, 320–325. https://doi.org/10.1016/j.neuropsychologia.2012.09.037.
Schwartze, M., Kotz, S.A., 2013. A dual-pathway neural architecture for specific temporal prediction. Neurosci. Biobehav. Rev. 37, 2587–2596. https://doi.org/10.1016/j.neubiorev.2013.08.005.
Sommer, M.A., Wurtz, R.H., 2008. Brain circuits for the internal monitoring of movements. Annu. Rev. Neurosci. 31, 317–338. https://doi.org/10.1146/annurev.neuro.31.060407.125627.
Sowman, P.F., Kuusik, A., Johnson, B.W., 2012. Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. Exp. Brain Res. 222, 149–157. https://doi.org/10.1007/s00221-012-3204-7.
Tian, X., Poeppel, D., 2015. Dynamics of self-monitoring and error detection in speech production: evidence from mental imagery and MEG. J. Cogn. Neurosci. 27, 352–364. https://doi.org/10.1162/jocn_a_00692.
Tian, X., Poeppel, D., 2012. Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. Front. Hum. Neurosci. 6, 314. https://doi.org/10.3389/fnhum.2012.00314.
Timm, J., SanMiguel, I., Saupe, K., Schröger, E., 2013. The N1-suppression effect for self-initiated sounds is independent of attention. BMC Neurosci. 14, 2. https://doi.org/10.1186/1471-2202-14-2.
Timm, J., Schönwiesner, M., Schröger, E., SanMiguel, I., 2016. Sensory suppression of brain responses to self-generated sounds is observed with and without the perception of agency. Cortex 80, 5–20. https://doi.org/10.1016/j.cortex.2016.03.018.
Van Petten, C., Luka, B.J., 2012. Prediction during language comprehension: benefits, costs, and ERP components. Int. J. Psychophysiol. 83, 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015.
Vercillo, T., O'Neil, S., Jiang, F., 2018. Action–effect contingency modulates the readiness potential. Neuroimage 183, 273–279. https://doi.org/10.1016/j.neuroimage.2018.08.028.
von Holst, E., 1954. Relations between the central Nervous System and the peripheral organs. Br. J. Anim. Behav. 2, 89–94. https://doi.org/10.1016/S0950-5601(54)80044-X.
Wacongne, C., Naccache, L., Bekinschtein, T., van Wassenhove, V., Labyt, E., Dehaene, S., 2011. Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc. Natl. Acad. Sci. 108, 20754–20759. https://doi.org/10.1073/pnas.1117807108.
Wang, J., Mathalon, D.H., Roach, B.J., Reilly, J., Keedy, S.K., Sweeney, J.A., Ford, J.M., 2014. Action planning and predictive coding when speaking. Neuroimage 91, 91–98. https://doi.org/10.1016/j.neuroimage.2014.01.003.
Wang, W., Datta, H., Sussman, E., 2005. The development of the length of the temporal window of integration for rapidly presented auditory information as indexed by MMN. Clin. Neurophysiol. 116, 1695–1706. https://doi.org/10.1016/j.clinph.2005.03.008.
White, P.M., Yee, C.M., 2006. P50 sensitivity to physical and psychological state influences. Psychophysiology 43, 320–328. https://doi.org/10.1111/j.1469-8986.2006.00408.x.
Whitford, T.J., Mathalon, D.H., Shenton, M.E., Roach, B.J., Bammer, R., Adcock, R. a, Bouix, S., Kubicki, M., De Siebenthal, J., Rausch, a C., Schneiderman, J.S., Ford, J.M., 2011. Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. Psychol. Med. 41, 959–969. https://doi.org/10.1017/S0033291710001376.
Widmann, A., Schröger, E., Maess, B., 2015. Digital filter design for electrophysiological data - a practical approach. J. Neurosci. Methods 250, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002.
Wolpert, D.M., Flanagan, J.R., 2001. Motor prediction. Curr. Biol. 11, R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8.
Woods, D.L., 1995. The component structure of the N1 human auditory evoked potential. Electroencephalogr. Clin. Neurophysiol. 44, 102–109.
Zouridakis, G., Simos, P.G., Papanicolaou, A.C., 1998. Multiple bilaterally asymmetric cortical sources account for the auditory N1m component. Brain Topogr. 10, 183–189. https://doi.org/10.1023/A:1022246825461.